School of Computing and Information Systems
The University of Melbourne
COMP90049
Knowledge Technologies (Semester 1, 2019)
Workshop exercises: Week 3

1. Following on from last week, write a **regular expression** which will:

   (a) Match a string according to whether it contains a price (like $20 or $0.99, but not 11.30 or 0$n1a).

   (b) Match a number in scientific E notation (e.g. `2.00600e+003`)

   (c) Remove all HTML comments from an HTML document (defined as a string)

   (d) Validate an email address (i.e. the string will match if it is an email address, and will mismatch otherwise)

&

---

2. Consider searching the string `muddle-the-middle-muddled-mud` for the query string `led-`:

   (a) How many comparisons are required for the "naive" approach?

   (b) Identify and construct the extra data structure required to use the version of the Boyer–Moore algorithm discussed in the lecture. (Note that the formal Boyer–Moore algorithm has a richer data structure.) How much extra space is required? How many comparisons within the search string are required? How many operations on the extra data structure?

3. Consider compressing the string `muddle-the-middle-muddled-mud`:

   (a) Show the dictionary that would be built using the simple form of LZ coding shown in lectures. Then show the final encoded string, using the lecture notation.